

1999

An Investigation of the Construct Validity of the Children's Skills Test

Cynthia D. L. Ladd

Eastern Illinois University

This research is a product of the graduate program in [School Psychology](#) at Eastern Illinois University. [Find out more](#) about the program.

Recommended Citation

Ladd, Cynthia D. L., "An Investigation of the Construct Validity of the Children's Skills Test" (1999). *Masters Theses*. 1677.
<https://thekeep.eiu.edu/theses/1677>

This is brought to you for free and open access by the Student Theses & Publications at The Keep. It has been accepted for inclusion in Masters Theses by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.

THESIS REPRODUCTION CERTIFICATE

TO: Graduate Degree Candidates (who have written formal theses)

SUBJECT: Permission to Reproduce Theses

The University Library is receiving a number of request from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow these to be copied.

PLEASE SIGN ONE OF THE FOLLOWING STATEMENTS:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university or the purpose of copying it for inclusion in that institution's library or research holdings.

Author's Signature

4-15-99
Date

I respectfully request Booth Library of Eastern Illinois University **NOT** allow my thesis to be reproduced because:

Author's Signature

Date

An Investigation of the Construct Validity

of the Children's Skills Test

(TITLE)

BY

Cynthia D. L. Ladd

1992 -

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Specialist in School Psychology (SSP)

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

1999

YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

9-20-99

DATE

9-20-99

DATE

ABSTRACT

The Children's Skills Test (CST) is an aptitude and achievement test for children in grades Kindergarten through Nine, developed from the American School Psycho-Educational Assessment Battery (ASPAB). It is designed for home administration via the computer, and no examiner is present. At this time, no technical information has been released regarding the CST, nor has any research been published.

The Kaufman Brief Intelligence Test (K-BIT) is a brief measure of general intelligence with sound technical properties. It is administered by an examiner. The K-BIT has been correlated with the Wechsler Intelligence Scale for Children, Third Edition (WISC-III), the most commonly used comprehensive measure of general intelligence.

This study examined the correlations between the CST, a new instrument, and the K-BIT, a well-researched instrument, to determine the construct validity of the CST-Aptitude Test. The magnitude of any significant correlations was examined, as were the standard score differences between the instruments.

Results indicated that CST composite, verbal, and nonverbal scores correlated significantly with the K-BIT scores. However, the magnitude of these correlations varied, and significant standard score differences were found between CST and K-BIT scores. In general, it was concluded that the CST was not suitable for use by parents or school psychologists at this time. This determination was made due to the lack of reliability and validity information available regarding the CST, and due to the significant score differences found between the CST and the K-BIT.

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr. J. Michael Havey, and my committee members, Dr. Gary Canivez and Dr. Christine McCormick, for their advice, guidance, and patience. I would also like to thank the Neoga, IL, school district, the staff, and the students for their invaluable participation in my research. Finally, I wish to thank my husband for his support and encouragement throughout this lengthy process.

INTRODUCTION

All standardized and norm-referenced tests must prove evidence of validity prior to their adoption for use by professional psychologists in any school system or agency. This is especially imperative for measures of general intelligence, as these instruments are frequently used to make major decisions about a child's educational future. It is necessary, but not sufficient, that an assessment instrument possess good reliability, have adequate norms, and be free from bias. Each instrument must also demonstrate validity.

The validity of a given instrument refers to the assessment's ability to measure the construct purported to be measured. For example, does the given intelligence test measure intelligence, or does it, perhaps, measure fluency in the English language? There are three aspects of validity: content, criterion-related, and construct. Every instrument must prove adequate in all three dimensions to be judged as valid.

Content validity is the extent to which individual test items appear to measure the given construct. Items must be appropriate for the construct, represent an adequate sampling of the domain, and be presented in an appropriate manner. Analysis of items can be done in three ways: face validity (does the item appear to address the correct domain), expert opinion, and statistical analysis. Typically, test developers use all three methods to ensure the instrument has adequate content validity. Without content validity, it is impossible to support a claim of test validity.

Criterion-related validity is the second dimension of validity. To demonstrate criterion-related validity, an instrument must show that it correlates to other criteria related to the construct. Instruments need to related to current criteria as well as predict future performance. For example, an intelligence test should correlate with performance

on academic achievement tests and should be able to predict future school performance. In short, criterion-related validity is necessary for the instrument to connect to “real world” situations. Without this generalizability, the instrument does not prove useful.

The final aspect of validity, construct validity, is essential for establishing overall validity. Construct validity refers to the instrument’s ability to measure the specific trait it is designed to measure. That is, does the score obtained from the instrument truly reflect the given construct or does it reflect a different, but related, construct. Face validity, or the surface appearance of suitability, is not sufficient to demonstrate construct validity. Typically, multiple research studies must be done, correlating subjects’ performance on the instrument with simultaneous performance on previously validated instruments designed to measure the same construct. For example, when the Kaufman Brief Intelligence Test (K-BIT) (Kaufman & Kaufman, 1990) was being validated, many studies were conducted correlating the subjects’ K-BIT scores to scores they obtained on established comprehensive measures of general intelligence, such as the Wechsler Intelligence Scale for Children, Third Edition (WISC-III) and the Stanford-Binet Intelligence Test, Fourth Edition (SB:IV). The WISC-III and SB:IV have long been the standards in the field of intelligence testing; without being able to demonstrate correlations with these two instruments, it is doubtful that the K-BIT would have been accepted as an adequate measure of intelligence.

REVIEW OF LITERATURE

The computer revolution has wrought changes throughout society. People no longer type, they keyboard; mail is often no longer sent through the post office, but via e-mail; business deals and political maneuvering are done on-line; and children research their papers using website databases. Almost all schools have computers for students to use, and more than 40% of homes had personal computers as of 1996 (World Almanac and Book of Facts, 1997). That figure is expected to reach near 60% by the year 2000 (Information Please, 1997). Not surprisingly, the educational software market is one of the most lucrative and profitable around (Annicelli, 1996; Hisey, 1996), outstripping the entertainment software market (Annicelli, 1996). A recent entry into this field is the Children's Skills Test (Virtual Knowledge, 1997a; Virtual Knowledge, 1997b; Virtual Knowledge, 1997c).

The Children's Skills Test (CST), published by Virtual Knowledge (1997a, 1997b, 1997c), is available in three versions: Primary (grades K-3), Intermediate (grades 4-6), and Advanced (grades 7-9). It is designed to assess a child's aptitude and achievement, via a computer-interface, and is marketed toward parents of school-aged children. No examiner or adult is present during the self-guided assessment session. Given the desire of today's parents to have input into their child's education, this program seems likely to be extremely popular with parents (Hickman, 1997). As such, it is necessary to demonstrate that the CST possesses adequate psychometrics to be considered a valid and reliable instrument.

The CST consists of two batteries, the Aptitude Test and the Achievement Test. For the purposes of this study, the focus will be on the Aptitude Test. According to the

Parent Handbooks (Virtual Knowledge, 1997a; Virtual Knowledge, 1997b; Virtual Knowledge, 1997c), the Aptitude Test measures general intelligence; it would best be characterized as a brief measure of general intelligence, as it only consists of either three or four subtests, depending on the child's grade level.

For students in grades Kindergarten through Three (ages 5.0 - 9.11), the Aptitude Test consists of three subtests: Listening Comprehension, Oral Directions, and Sequencing. According to the Parent Handbook (Virtual Knowledge, 1997a), the Listening Comprehension subtest "measures receptive vocabulary, the ability to understand the meaning of words that have been heard or read" (p. 16). The Oral Directions subtest "measures . . . ability to follow simple or complex instructions spoken aloud. It includes elements known as short-term sequential memory and vocabulary" (p. 16). The final aptitude subtest, Sequencing, "measures abstract abilities of a nonverbal nature, specifically the ability to recognize sequential relationships" (p. 18). The raw scores from the subtests are converted to scaled scores for each subtest, and a general ability score (GAQ), based on all three subtests, is also reported.

For children in grades Four through Nine (ages 9.0 - 15.11), the Aptitude Test is composed of four subtests: Verbal Analogies, Listening Vocabulary, Symbolic Reasoning, and Figural Analogies (Virtual Knowledge, 1997b; Virtual Knowledge, 1997c). According to the Parent Handbooks (Virtual Knowledge, 1997b; Virtual Knowledge, 1997c), the Verbal Analogies subtest "measures the ability to recognize relationships of material presented aloud" (p. 15) and the Listening Vocabulary subtest "measures vocabulary proficiency" by having children determine "if two words mean the same thing, the opposite, or neither" (p. 15). The Symbolic Reasoning subtest measures nonverbal

reasoning via matrices and the Figural Analogies subtest measures the child's ability to choose the picture that best completes a pictorial analogy.

The Parent Handbooks (Virtual Knowledge, 1997a; Virtual Knowledge, 1997b; Virtual Knowledge, 1997c) report that the CST is a computer-based revision of the American School Psycho-Educational Assessment Battery (ASPAB) (Bryant, Mathews, Ammer, Cronin, Mandelbaum, & Quinby, date unknown), which consists of both the American School Achievement Test and the American School Aptitude Scale. The ASPAB is a group-administered pencil-and-paper test of aptitude and achievement, for students in grades Kindergarten through Nine. The Handbooks report that form-equivalence studies were done between the ASPAB and the CST, and that the "tests are essentially interchangeable" (p. 52). Additionally, the Handbooks indicate that, since the CST and ASPAB are equivalent, the statistical properties of the ASPAB apply to the CST. The Handbooks, however, do not go on to detail the nature of those statistical properties or provide any of the statistics for scrutiny. No documentation of the technical characteristics of the CST is provided in the Handbooks.

The Parent Handbooks (Virtual Knowledge, 1997a; Virtual Knowledge, 1997b; Virtual Knowledge, 1997c) do report that additional research has been done on the technical adequacy of the CST itself. For example, the pool of original ASPAB items plus items new to the CST were subjected to expert opinion and statistical analyses, to remove biased and unreliable items. The final items were field-tested across the country with various children, to validate item-order. The Handbooks also reported that the CST was normed on a sample based on the US census and included "sufficient numbers . . . representative of the nation as a whole" (p. 50). Reliability, face validity, and concurrent

validity are reported by Handbooks, as is the CST's ability to discriminate between low and high achievers.

However, as noted before, no documentation of the research establishing reliability or validity is provided in the Parent Handbooks (Virtual Knowledge, 1997a; Virtual Knowledge, 1997b; Virtual Knowledge, 1997c) or in published research studies. Nor are any of the actual statistical data pertaining to the norm sample, the reliability, or the validity of the CST are provided in the Parent Handbooks. Therefore, it is unknown whether the standardization sample was adequate, whether the test has good internal consistency and test-retest reliability, or whether the CST has demonstrated criterion-related or construct validity. The Handbooks refer interested parents to a separately available Technical Manual for more information regarding psychometrics. However, when contacted by phone, a representative of Virtual Knowledge indicated that the Technical Manual is not yet available (personal communication, 12/97).

Virtual Knowledge did provide, upon request, a photocopy of the Examiner's Technical and Normative Manual (Bryant et al., date unknown), omitting the copyright and publisher information pages, for the most recent edition of the ASPAB. As noted previously, the Parent Handbooks (Virtual Knowledge, 1997a; Virtual Knowledge, 1997b; Virtual Knowledge, 1997c) report that the ASPAB and CST are technically identical.

According to the information provided in this Technical Manual, the ASPAB has a mean of 100 and a standard deviation of 15. In general, the manual reported adequate norms, nearly adequate reliability, and generally adequate validity. However, this information cannot be relied upon, as the Parent Handbooks of the CST (Virtual Knowledge, 1997a; Virtual Knowledge, 1997b; Virtual Knowledge, 1997c) imply that

new items have been added to the ASPAB during its revision to the CST. The addition of new items would necessitate new reliability and validity studies, not a reliance on technical data from the ASPAB. Additionally, the Parent Handbooks state that form equivalence studies have been conducted correlating the paper-and-pencil ASPAB to the computerized CST. However, these results are not yet available to the public. Despite the information supplied about the ASPAB, it is obvious that current information regarding normative data, reliability, and validity needs to be supplied specifically about the CST. It cannot be assumed, despite assertions in the Handbooks, that the CST and ASPAB share the same technical properties.

As stated above, many psychometric aspects of the CST need to be investigated. One of the most important measures of validity that an instrument must meet is that of convergent validity, which is an aspect of construct validity. That is, does the CST correlate positively and highly with other measures of aptitude and achievement? The Handbooks report that the CST correlates well with other measures of aptitude; however, no data are provided, no correlated instruments are listed, and no outside research has been done at this date. Therefore, it must be determined whether the CST-Aptitude Test correlates highly with another respected measure of general intelligence.

The Kaufman Brief intelligence Test (K-BIT) is an individually administered brief screening measure of intelligence designed for use with people from ages 4 to 90 (Kaufman & Kaufman, 1990; Parker, 1993; Hildman, Friedberg, & Wright, 1993). Various reviews have concluded that the K-BIT is a psychometrically sound instrument that is ideally suited to the estimation of general intelligence when one cannot, or need

not, administer a comprehensive intelligence scale (Hildman et al., 1993; Parker, 1993; Canivez, 1995; Prewett, 1995).

According to the manual (Kaufman & Kaufman, 1990), the K-BIT takes 15 to 30 minutes to administer and is divided into two main sections, Vocabulary and Matrices, designed to assess verbal and nonverbal aptitude. The Vocabulary section is further divided into two subtests, Expressive Vocabulary and Definitions. These subtests involve naming a pictured object and producing a word given two clues, respectively. The Matrices section is a multiple-choice task that requires the examinee to identify the correct completion to the matrix based on understanding relationships between various visual stimuli.

The K-BIT has a mean of 100 and a standard deviation of 15, similar to other measures of intelligence. It was normed on a sample of 2022 adults and children, following the 1990 census projections. The sample was stratified by gender, geographic region, socioeconomic status, and ethnicity. The standardization sample for the K-BIT deviated slightly from the census figures, but at levels much less than the ASPAB.

According to the K-BIT manual (Kaufman & Kaufman, 1990), the internal consistency reliability of the K-BIT as a whole is reported as .93, averaged across all age ranges. Test-retest reliability for the entire test averaged .94 across all age ranges. Evidence of concurrent and construct validity was also reported. Concurrent validity was found with the Kaufman Assessment Battery for Children (K-ABC) and the Wechsler Intelligence Scale for Children-Revised (WISC-R). The K-BIT correlated .63 with the K-ABC and .80 with the WISC-R. In demonstration of construct validity, it was shown that the K-BIT's verbal/nonverbal divisions correlated with those in other instruments; that is,

the Vocabulary section correlated more highly with the WISC-R and WAIS-R Verbal IQs than it did with Performance IQs.

In addition to evidence presented in the manual, a number of research studies have demonstrated that the K-BIT possesses adequate reliability and validity and, furthermore, that it correlates well with other, more comprehensive, measures of intelligence (see Table 1). Donovan, Burright, Burg, Gronendyke, Klimczak, Matthews, and Sardo (1996) administered the K-BIT to a diverse group of 196 people and found support for the Manual's claim that the K-BIT is, indeed, useful for estimating ability across populations and clinical groups.

When studied using samples of young (preschool and kindergarten) students, the K-BIT correlated .75 with the Peabody Picture Vocabulary Test-Revised (PPVT-R) and there was 83% agreement between instruments when classifying children as average or above average (Childers, Durham, & Wilson, 1994). In a separate study, the K-BIT was found to correlate .63 with the Wechsler Preschool and Primary Scales of Intelligence-Revised (WPPSI-R), with students earning an average of 2.1 points higher on the K-BIT (Lassiter & Bardos, 1995).

For older individuals, the most commonly used measures of intelligence are the Wechsler scales (Sattler, 1988). When the K-BIT was published, the Wechsler Intelligence Scale for Children-Revised (WISC-R) was in use. Prewett (1992a) reported a correlation of .81 between the WISC-R and the K-BIT for students (mean age 10.8) referred for psychoeducational assessment. However, the K-BIT scores averaged six points lower than the WISC-R scores. In another study, it was found that the K-BIT correlated .64 with the WISC-R when administered to incarcerated male juvenile

delinquents (mean age 15.9) (Prewett, 1992b). In this study, there was a 0.45 point difference between the K-BIT and the WISC-R scores, with WISC-R scores being higher.

In a longitudinal study with 44 adolescents, a correlation of .80 was found between the WISC-R and K-BIT (Slate, Graham, & Bower, 1996). In this study, children had been administered the WISC-R three years prior to the administration of the K-BIT. An average score difference of 1.8, in favor of the WISC-R, was found between the instruments. This supports Prewett's study of juvenile delinquents (1992b), but is in contrast to the results obtained in his study of students referred for psychoeducational evaluation (1992a). However, these three studies do demonstrate that there is a high correlation between scores on the WISC-R and scores on the K-BIT, regardless of which score is higher. These studies help establish the concurrent validity of the K-BIT when the WISC-R is used as the standard.

However, WISC-R was revised in 1992 and the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) is now the most commonly administered measure of general intelligence (Stinnett, Havey, & Oehler-Stinnett, 1994). In a study of 29 students (mean age 7.96), referred for gifted evaluation, the K-BIT was found to correlate .53 with the WISC-III, with WISC-III scores averaging 2.27 higher (Levinson & Folino, 1994a; Levinson & Folino, 1994b), which demonstrates concurrent validity. Prewett (1995) found a correlation of .78 between the K-BIT and WISC-III in a study of 50 students referred for psychoeducational evaluation (mean age 9.3). In this study, K-BIT scores averaged 4.8 points higher than did WISC-III scores. In a larger study of referred students (Canivez, 1995), it was found that the K-BIT correlated with the WISC-III .87

for a group of 137 students (mean age 11.7), with an average score difference of .73 in favor of the K-BIT.

When examined for use with students identified with disabilities, the K-BIT has also demonstrated concurrent validity with the WISC-III. In a study of 75 students identified as learning disabled (mean age 11.79), it was found that the K-BIT correlated .82 with the WISC-III, with an average score difference of 1.67 in favor of the WISC-III (Canivez, 1996; Canivez, 1996). Jovorsky (1993) investigated the use of the K-BIT with 63 children (mean age 13.9) who had been identified with psychiatric disorders. For these children, it was found that the K-BIT correlated .76 with the WISC-III, with an average score difference of 1.61 in favor of the WISC-III.

The Stanford-Binet: Fourth Edition (SB:IV) is another comprehensive measure of general intelligence that is often used when assessing students (Stinnett et al., 1994). Prewett and McCaffery (1993) found that the K-BIT correlated .81 with the SB:IV, with students earning an average of 5.1 points higher on the SB:IV. When examined in conjunction with another instrument used by the public schools, the Slosson Intelligence Test (SIT), it was found that the K-BIT correlated an average .67 with the SIT (Hildman et al., 1993). Overall, only one study was examined where the K-BIT failed to correlate $\geq .60$ with an established instrument (Levison & Folino, 1994a). Taken in total, these studies provide ample evidence that the K-BIT correlates with the WISC-III, the most commonly used and psychometrically sound measure of general intelligence used with the school-aged population (Salvia & Ysseldyke, 1995), and with other instruments used with school-aged children.

Table 1
Correlations between the K-BIT and other instruments

Study	Instrument	Correlation	Ave. point Difference
Childers et al., 1994	PPVT-R	.75	---
Lassiter & Bardos, 1995	WPPSI-R	.63	2.1
Kaufman & Kaufman, 1990	K-ABC	.63	---
Kaufman & Kaufman, 1990	WISC-R	.80	---
Prewett, 1992a	WISC-R	.81	(6.0)
Prewett, 1992b	WISC-R	.64	(0.45)
Slate et al., 1996	WISC-R	.80	(1.8)
Levinson & Folino, 1994a	WISC-III	.53	(2.27)
Prewett, 1995	WISC-III	.78	4.8
Canivez, 1995	WISC-III	.87	0.73
Canivez, 1996	WISC-III	.82	(1.67)
Jovorsky, 1993	WISC-III	.76	(1.61)

(n) = K-BIT average score was lower than the comparison
instrument's average score

PURPOSE

The purpose of this study was to determine if the Aptitude Test of the Children's Skills Test (CST) correlates to a degree sufficient to demonstrate adequate construct validity with an established measure of general intelligence. The Kaufman Brief Intelligence Test (K-BIT) has demonstrated consistent correlations with the standard in intelligence testing, the Wechsler Intelligence Scale for Children, Third Edition (WISC-III) and other established measures of general intelligence. Additionally, the K-BIT only takes 15-30 minutes to administer, whereas the WISC-III can take up to 90 minutes to administer; this makes the K-BIT a more efficient research tool. Finally, the K-BIT is a brief measure of general intelligence, as is the Aptitude Test of the CST, as each instrument contains three to four subtests. Therefore, the K-BIT is a logical instrument to use as the comparison for the CST. Although other aspects of the CST's reliability and validity also need to be investigated, the focus of this current study will be limited to an investigation of the CST's construct validity.

RESEARCH QUESTIONS

This study attempted to answer the following research questions:

1. Do the composite, verbal, and nonverbal scores from the Aptitude Test of the Children's Skills Test, a computer-based and individually self-administered test of aptitude, correlate significantly with the composite, vocabulary, and matrices scores of the Kaufman Brief Intelligence Test, an individually examiner-administered brief test of general intelligence?
2. Are these correlations $+ .60$ or greater (the level suggested by Sattler, 1988) in magnitude?
3. Is there a significant ($p < .01$) difference between K-BIT scores and CST scores (composite, verbal, and nonverbal)?

METHOD

Recruitment of Subjects

A letter was sent to 4th, 5th, and 6th grade teachers (Appendix A) and interested teachers were then provided with parent consent forms and sent them home (Appendix B). Interested parents returned the consent forms to the teachers. The researcher collected the consent forms from the teachers. Arrangements were made with the classroom teacher so that the student missed minimal academic time. Some students arranged for testing outside of school hours.

Measures

Children's Skills Test (Virtual Knowledge, 1997a; Virtual Knowledge, 1997b; Virtual Knowledge, 1997c). The CST is an individual test of aptitude and achievement administered via computer, designed for students in grades kindergarten through nine. The Aptitude Battery takes between 40 and 60 minutes to complete, and consists of four subtests: verbal analogies, listening vocabulary, symbolic reasoning, and figural analogies. A Verbal Composite is calculated from the verbal analogies and listening vocabulary subtests, and a Nonverbal Composite is derived from the symbolic reasoning and figural analogies subtests. In addition to the composites, an overall IQ score (GAQ) is also reported. No specific information is available at this time about standardization, reliability, or validity.

Kaufman Brief Intelligence Test (Kaufman & Kaufman, 1990). The K-BIT is an individually administered brief measure of general intelligence, designed to be administered to people ages 4 to 90. The K-BIT takes between 15 and 30 minutes to administer and consists of two sections, Vocabulary and Matrices; Vocabulary contains two subtests,

Expressive Vocabulary and Definitions. Internal consistency estimates average .94, as do test-retest reliability estimates. The K-BIT was standardized on a representative sample reflecting 1990 census estimates.

Procedure

First, subjects were recruited (see previous section), and ID numbers were assigned. The subjects were administered the K-BIT and CST-Aptitude Battery in counterbalanced order. During the CST administration, the students did not interact with the examiner as per the procedure stated in the Handbooks. The subjects took the CST on a laptop computer equipped with a CD-ROM, speaker, and a color monitor, or a comparably equipped desk-top computer. The K-BIT was administered to the subjects by the examiner and required one protocol per person. The assessment session took approximately one hour and fifteen minutes and took place in the counseling office of an elementary school, or an office at a local university.

Information was coded by ID number for the purposes of confidentiality.

Subjects

The sample population in this study consisted of thirty elementary school students in grades four through six. Ninety percent of the students attended a rural school district, with approximately 830 students. The other ten percent attended a larger school district with approximately 3400 students; both districts were located in east central Illinois. Of the students in the sample, forty-three percent were female and fifty-seven percent male. Forty percent of the students were in fourth grade, thirty-seven percent in fifth grade, and twenty-three percent were in sixth grade. Thirty percent of the students in the sample had previously been diagnosed with learning disabilities and were receiving special education

services. All of the students in the sample were Caucasian, which reflected the ethnic composition of the geographic area.

The thirty subjects were administered the Kaufman Brief Intelligence Test (K-BIT) and the Children's Skills Test-Aptitude Test (CST) in counterbalanced order. Students with diagnosed motor difficulties (e.g., receiving occupational or physical therapy services) were excluded from the study, due to the obvious confounding difficulties associated with the use of a computer mouse. Additionally, students whose primary language was not English were excluded, as the CST is only available in English.

Table 2
Demographic Characteristics

Characteristic	% of sample	N
GENDER		
Male	57%	17
Female	43%	13
GRADE		
Grade 4	40%	12
Grade 5	37%	11
Grade 6	23%	7
AGE		
Age 9	10%	3
Age 10	23%	7
Age 11	43%	13
Age 12	23%	7
DISABILITY		
Learning Disability	33%	10
No eligibility	67%	20
SCHOOL DISTRICT		
Rural School District	90%	27
Larger School District	10%	3

Statistics

Pearson's correlations were calculated to determine the correlations among composite, verbal/vocabulary, and nonverbal/matrices scores. Paired t-tests (for

dependent measures) were conducted to determine the average score differences between tests and the significance of those differences. All statistical analyses were run on SPSS 7.5, using a Windows 95 platform.

RESULTS

This study was designed answer three simple research questions. (1) Do the composite and subscale scores from the Aptitude Battery of the CST correlate significantly with the composite and subscale scores of the K-BIT, (2) Are those correlations of a magnitude of $+ .60$ or greater, and (3) Is there a significant ($p < .01$) difference between CST and K-BIT composite and subscale scores. To answer these questions, Pearson's correlations and paired t-tests (for dependent measures) were conducted.

For the research study sample, the composite scores of the CST and K-BIT do, indeed, correlate significantly ($r = .74$, $p < .000$). A significant correlation ($r = .73$, $p < .000$) was also found between CST verbal scores and K-BIT vocabulary scores, and between CST nonverbal and the K-BIT matrices score ($r = .59$, $P < .001$) (Table 4).

The correlation between the composite scores of the CST and the K-BIT did, indeed, meet the magnitude standard set forth in the research questions ($\geq +.60$). Similarly, so did the correlation between the CST verbal and K-BIT vocabulary scale scores. Although significant, the correlation between CST nonverbal and K-BIT matrices scores did not meet the magnitude standard. However, that correlation of $.59$ was extremely close to the criterion of $.60$. Therefore, it can be said that as CST composite scores rose, so did K-BIT composite scores. Similarly, higher scores on the verbal portion of the CST correlated to higher scores on the vocabulary section of the K-BIT, and higher scores on the nonverbal portion of the CST correlated to higher scores on the K-BIT.

Paired t-tests were calculated to determine if significant differences between the CST and K-BIT composite and subscale scores existed. There was an average difference

of 7.13 points between the CST and K-BIT composite scores, with the K-BIT scores higher. This difference was determined to be statistically significant ($p < .001$). This also held true for the verbal and vocabulary portions of the CST and K-BIT. K-BIT vocabulary scores averaged 12.43 points higher than CST verbal scores, and this difference was statistically significant ($p < .000$). the CST nonverbal and K-BIT matrices scores only differed an average of 0.13 points, and that difference was not found to be significant ($p < .954$).

For the research sample, the composite scores of the K-BIT and CST correlated significantly. However, there was a significant difference between the resulting composite IQ scores. Similarly, the vocabulary and verbal scores of the K-BIT and CST also correlated and also demonstrated a significant difference in subscale scores. The matrices and nonverbal subscale scores of the K-BIT and CST correlated significantly, but not to the magnitude required, and the scores did not differ significantly. Therefore, it seems that the composite, verbal, and nonverbal CST scores correlate well with the K-BIT composite, vocabulary, and matrices scores for the sample studied, but that composite and verbal scores on the CST are significantly lower than on the K-BIT.

Table 3
Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Age	10.60	.97	9	12
K-BIT composite	97.97	13.64	58	117
K-BIT vocabulary	99.83	12.48	73	119
K-BIT matrices	96.13	14.15	50	115
CST composite	90.83	16.22	65	117
CST verbal	87.40	17.46	61	121
CST nonverbal	96.00	13.38	73	121

N=30

Table 4
Correlations and Significance

Paired Samples	Correlation	Sig.	Difference b/w Means	T-test Sig. (2-tailed)
CST composite & K-BIT composite	.74	.000	7.13	.001
CST verbal & K-BIT vocabulary	.73	.000	12.43	.000
CST nonverbal & K-BIT matrices	.59	.001	0.13	.954
N=30				

DISCUSSION

The Children's Skills Test (CST) affords an opportunity for parents to assess and monitor the academic progress of the children. Virtual Knowledge, the test developer, indicates in the Parent Handbooks (Virtual Knowledge, 1997a; Virtual Knowledge, 1997b; Virtual Knowledge, 1997c) that the test is as reliable and valid as test results provided by the school. The results of this study indicate the overall CST aptitude battery scores do correlate with an established measure of intelligence, specifically the K-BIT, as stated in the Parent Manual, for the research sample. However, CST composite scores are significantly lower than those resulting from the K-BIT.

There can be many possible hypotheses regarding these score differences. It is speculated that these lower scores may be the result of the CST's presentation format. Unlike the K-BIT, all aspects of the CST are timed, which adds an inherent measure of stress to the session that could impact the student. Timing also limits the opportunity for the student to recall information and formulate responses. Additionally, students cannot ask to have information repeated and cannot take a break if needed. Finally, there is no examiner present during the CST assessment, whereas the K-BIT is presented by a trained evaluator. It is possible that the lack of human interaction, social cues, and rapport could also impact the CST scores.

There could also be significant differences in the normative samples used to standardize the assessments. For example, the K-BIT used parent education level as a measure of socio-economic status, and the CST used family income. Therefore, quite different populations could have been included in the norm groups. It could also be possible that the K-BIT scores are inflated, either due to norm drift or "easier" items.

However, it is not possible to determine if the type or age of the CST and K-BIT norms differ significantly, as the researcher has been unable to determine when or how the CST was standardized. Additionally, previous research has indicated that the K-BIT correlates well with established comprehensive measures of general intelligence and that significant score differences were not noted (Table 1).

Score differences could also have occurred if the two assessments were measuring different forms of the general constructs of verbal and nonverbal reasoning. In the verbal area, the K-BIT assesses concrete and abstract vocabulary development. The CST assesses synonyms, antonyms, and analogies. Taken at face value, it would seem that these subtests essentially tap into different areas of verbal reasoning skills. However, the nonverbal portion of the CST does have matrix reasoning (the symbolic reasoning subtest), very similar to that on the K-BIT. The CST then adds figural analogies, a task not represented on the K-BIT. Therefore, score differences in the verbal domain and weaker correlations in the nonverbal domain may be due to the assessment of different constructs.

Due to the limited number of students in this research sample and the significant scores differences between instruments, it is obvious that more research must be conducted to validate the CST Aptitude Test. A large, representative sample would be beneficial, as would specific research regarding students with disabilities and minority children. It is necessary to determine if the pattern of scores found in this study is replicated in future research. As mentioned previously, standard reliability research also needs to be conducted, as do further validity studies.

Overall, findings seem to indicate that the current utility of the Aptitude Battery of the Children's Skills Test is limited. It correlated adequately with an established measure of general intelligence and can yield a rough, although low, estimate of intellectual functioning. Unfortunately, this information will not typically be useful to parents, the target audience of the CST, as scores do not indicate areas of academic weakness or remediation strategies. As a significant discrepancy between CST and K-BIT scores exists, this information would also not be diagnostically useful to psychologists, a potential audience for the CST. Finally, the lack of reliability and validity evidence precludes the ethical use of the CST by assessment professionals.

REFERENCES

- Annicelli, C. (1996). Software sees smart growth. Playthings, 94(12), 38-39.
- Bryant, B. R., Mathews, S. C., Ammer, J. J., Cronin, M. E., Mandelbaum, L. H., & Quinby, S. S. (date unknown). American School Psycho-Educational Assessment Batteries: Examiner's technical and normative manual. [Information unavailable regarding copyright date and page numbers.]
- Canivez, G. L. (1995). Validity of the Kaufman Brief Intelligence Tests: Comparisons with the Wechsler Intelligence Scale for Children-Third Edition. Assessment, 2, 101-111.
- Canivez, G. L. (1996a). Validity and diagnostic efficiency of the Kaufman Brief Intelligence Test in reevaluating students with learning disability. Journal of Psychoeducational Assessment, 14, 4-19.
- Childers, J. S., Durham, T. W., & Wilson, S. (1994). Relation of performance on the Kaufman Brief Intelligence Test with the Peabody Picture Vocabulary Test - Revised among preschool children. Perceptual and Motor Skills, 79, 1195-1199.
- Donovick, P. J., Burright, R. G., Burg, J. S., Gronendyke, S. J., Klimczak, N., Matthews, A., & Sardo, J. (1996). The K-BIT: A screen for IQ in six diverse populations. Journal of Clinical Psychology in Medical Settings, 3, 131-139.
- Hickman, A. (1997). Children's IQ and Achievement Test. PC Magazine, 16(15), 367.
- Hildman, L. K., Friedberg, P. M., & Wright, P. M. (1993). Review of the Kaufman Brief Intelligence Test. Journal of Psychoeducational Assessment, 11, 98-101.

Hisey, P. (1996). Software for kids grows up. Discount Store News, 35(1), 26-29.

Information Please Almanac, 1997. (1996). New York: Houghton-Mifflin

Company.

Javorsky, J. (1993). The relationship between the Kaufman Brief Intelligence Test and the Wechsler Intelligence Scale for Children-III in a clinical sample. Diagnostic, 19, 337-385.

Kaufman, A. S., & Kaufman, N. L. (1990). Kaufman Brief Intelligence Test. Circle Pines, MN: American Guidance Service.

Lassiter, K. S., & Bardos, A. N. (1995). The relationship between young children's academic achievement and measures of intelligence. Psychology in the Schools, 32, 170-177.

Levinson, E. M., & Folino, L. (1994a). Correlations of scores of the Gifted Evaluation Scale with those on the WISC-III and Kaufman Brief Intelligence Test for students referred for gifted evaluation. Psychological Reports, 74, 419-424.

Levinson, E. M., & Folino, L. (1994b). The relationship between the WISC-III and the Kaufman Brief Intelligence Test with students referred for gifted evaluation. Special Services in the Schools, 8, 155-159.

Parker, L. D. (1993). The Kaufman Brief Intelligence Test: An introduction and review. Measurement and Evaluation in Counseling and Development, 26, 152-155.

Prewett, P. N. (1992a). The relationship between the Kaufman Brief Intelligence Test (K-BIT) and the WISC-R with referred students. Psychology in the Schools, 29, 25-27.

Prewett, P. N. (1992b). The relationship between the Kaufman Brief Intelligence Test (K-BIT) and the WISC-R with incarcerated juvenile delinquents. Educational and Psychological Measurement, 52, 977-982.

Prewett, P. N. (1995). A comparison of two screening tests (the Matrix Analogies Test-Short Form and the Kaufman Brief Intelligence Test) with the WISC-III. Psychological Assessment, 7, 69-72.

Prewett, P. N., & McCaffery, L. K. (1993). A comparison of the Kaufman Brief Intelligence Test (C-BIT) and the Stanford-Binet, a two-subtest short form, and the Kaufman Test of Educational Achievement (K-TEA) Brief Form. Psychology in the Schools, 30, 299-304.

Sattler, J. (1988). Assessment of Children. San Diego, CA: Author.

Slate, J. R., Graham, L. S., & Bower, J. (1996). Relationships of the WISC-R and K-BIT for an adolescent clinical sample. Adolescence, 31, 777-782.

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. Journal of Psychoeducational Assessment, 12, 331-350.

Virtual Knowledge. (1997a). Children's Skills Test Handbook, Grades K-3. Needham, MA: Publisher.

Virtual Knowledge. (1997b). Children's Skills Test Handbook, Grades 4-6. Needham, MA: Publisher.

Virtual Knowledge. (1997c). Children's Skills Test Handbook, Grades 7-9. Needham, MA: Publisher.

World Almanac and Book of Facts, 1997. (1996). New Jersey: World Almanac Books.

APPENDIX A: LETTER TO TEACHERS

Dear Teachers,

I am a graduate student at Eastern Illinois University conducting research about computer-based testing. The test I am investigating, The Children's Skills Test, is designed for students in grades Kindergarten through Nine. I would like to have your assistance in recruiting children to participate in this project. Please see the attached parental consent form for details.

If you have any questions about this research project, you can contact me at 235-2507. You may also contact my supervisor, Dr. J. Michael Havey, at 581-2127. I will check back with you next week to collect the returned parental consent forms and answer any further questions you may have.

Thank you very much for your time. I look forward to working with you and your class members!

Sincerely,

Cindy Ladd, Graduate Student
Eastern Illinois University

APPENDIX B: PARENTAL CONSENT FORM & DEMOGRAPHIC SURVEY**Parental Consent & Survey
Description of Research Project**

Dear Parent(s):

Thank you very much for allowing your child to participate in this research project! This research project is investigating a new, computer-based test, called the Children's Skills Test. This computer test is designed to test a child's aptitude (or ability) in the areas of non-verbal and verbal learning. The Kaufman Brief Intelligence Test is the test to which the Children's Skills Test will be compared. This also tests verbal and non-verbal aptitude.

Your child's participation will involve taking both tests during one session. It is anticipated that this will take approximately sixty minutes. Your child will be assigned an ID number, and all information will be kept confidential. Your child's name will not be recorded on my data forms, and he or she will not be identified in the study.

As the Children's Skills Test is a new instrument, it has not yet been researched. Therefore, we don't know how good a test it is. For this reason, I will not be telling children or their parents what score they get on the test. However, when the research is complete, you may obtain a copy of the report by calling me at (217) 235-2507 or by writing to me at: Cindy Ladd, Department of Psychology, Eastern Illinois University, Charleston, IL 61920. Please contact me if you have any questions you wish answered prior to completing the consent form on the reverse side of this page.

Once again, thank you very much!

Sincerely,

Cindy Ladd, Graduate Student
Eastern Illinois University

Consent Form

I give consent for my child, _____, to participate in the described research project. I understand that my child's name will not be revealed and that all information will be kept confidential. I understand that my child's participation will take approximately 60 minutes, and that this participation can take place at my child's school or at Eastern Illinois University, based on my preference. I further understand that I have the right to rescind this consent form at any time, or have any questions answered, by contacting Cindy Ladd (235-2507) or Dr. J. Michael Havey (581-2127).

Signature of Legal Guardian

Date

Survey

Child's Date of Birth: _____ Grade Level: _____

Ethnic Background: _____ City of Residence: _____

Parent Education Level (circle highest level completed):

8th grade or less

Some HS

HS degree

Some college/technical school

College degree

Graduate school